# Forms of Artificial Neural Networks for International Crude Oil Price Forecasting Research Based on ADF and GCT Post-Tests

Junjie Zhao *, Guangjie Liu

School of Mechanical and Power Engineering, Shenyang University of Chemical Technology, Shenyang, China, 110142

* Corresponding Author Email: z2498107840@163.com

**Abstract.** Crude oil is the core of all types of energy for industrial production and transportation as well as for daily life, and is an important strategic resource relative to social security, world peace and the maintenance of its international status. With the changes in international crude oil prices and market finance and economy gradually began to closely linked between its price fluctuations not only affect the cost of energy and economic growth, but also on the financial market and international trade has a far-reaching impact. Firstly, the ADF test is used to determine the smoothness of the time series of crude oil prices; secondly, the causal relationship between variables with different time lags is analyzed by Granger causality test; finally, based on the results of the above test, the ANN forecasting model is constructed and optimized to forecast the international crude oil prices in the short term. The ANN forecasting model based on ADF and Granger test proposed in this study provides a new idea and method for international crude oil price forecasting, which has high practical application value.

**Keywords:** ADF test, Mechanical learning, Artificial neural networks, Crude oil price forecasting model.

## 1. Introduction

Traditional econometric and statistical models are difficult to capture the complex nonlinear features hidden in the oil market, and the development of artificial intelligence and machine learning models in the field of oil forecasting is immeasurable. Zhao et al. used a connectivity network model to analyze the risk spillovers between WTI returns and eight important financial factors. The final results showed that the total connectivity showed time-varying characteristics and increased dramatically during the financial crisis flag [1]. Li M et al. By combining the trajectory similarity method with a machine learning model based on the decomposition-integration framework and then predicting crude oil prices the proposed method performs best in different evaluation metrics and statistical tests at different levels, indicating that the proposed VMD-SE-TS/LSTM learning paradigm is effective and robust in crude oil price prediction. [2]. However, with the increase in hidden layers and hidden neurons, the increase in model complexity does not necessarily lead to higher nonlinear modeling accuracy. Nalini Gupta et al. Artificial Neural Networks (ANN) for predicting crude oil prices. The main advantage of this artificial neural network approach is that it consistently captures the unstable patterns of crude oil prices and controls the delayed effects of crude oil prices by finding out the optimal lags and quantities. lags over time, and then, the results can be verified by evaluating the root mean square error [3]. Witold Orzeszko et al. Based on a nonlinear Granger causality test, we find a strong bi-directional causality between the price of crude oil and the two currency pairs: there is a stronger bi-directional causality between the EUR/USD and GBP/USD, while there is a weaker causality between the price of crude oil and the JPY/USD. The causality between the crude oil price and JPY/USD is weak [4].

In this paper, we propose one on analyzing the research results of scholars at home and abroad, and after screening the factors affecting crude oil prices based on ADF and GCT models, each of them generates a different subset of indicators, which are incorporated into Artificial Neural Networks (ANN), Ridge Regression (RR), Support Vector Regression (SVR), Random Forests (RF), and Long and Short-Term Memory Networks (LSTM) machine learning algorithms respectively, to construct the crude oil price portfolio forecasting model. The mean square error (MSE) and mean absolute error (MAE) were used as loss functions, and the adjusted $R^2$ was used as the performance index, and the two were considered together to evaluate the model performance. The empirical results surface that the artificial neural network has the best prediction performance, with the smallest MSE and MAE values and an adjusted $R^2$ value of 0.9607. This result verifies the efficiency of the two-stage model screening in the previous section, and also realizes the correct prediction of the crude oil market pattern.

## 2. Rationale and methodology

### 2.1. ADF unit root test

Determine the test model: according to the characteristics of the time series, choose different test models. There are three main models: the model without constant term and trend term, the model with only constant term, and the model with constant term and linear trend term.

Selection of lag order p: The selection of lag order p has a great influence on the results of ADF test. Usually, the information criterion method (e.g. AIC, BIC, etc.) or the F-statistic method can be used to determine the optimal lag order p. The lag order p can also be reasonably selected according to the actual situation and relevant theoretical knowledge.

Calculating the ADF statistic: The series is first-order differenced to create an autoregressive model that describes the relationship between the amount of change and the amount of previous change. The model parameters are then estimated and the residual term is computed, and the residual term is used to compute the ADF statistic.

Hypothesis testing: The original hypothesis of the ADF test is that “there is a unit root in the series” and the alternative hypothesis is that “the series is smooth”. At a given confidence level (usually 95% or 99%), if the P-value of the ADF statistic is less than the confidence level, the original hypothesis is rejected and the series is considered not to have a unit root; otherwise, the original hypothesis cannot be rejected and the series is considered to have a unit root.

### 2.2. Granger causality test

The basic logic of the Granger Causality Test (GCT) is to compare the predictive power of two time series models: one model includes only the historical values of the target variable, while the other model includes both the historical values of the target variable and the potential causal variable. If the model's ability to predict the target variable is significantly improved after including the historical values of the potential causal variable, the potential causal variable is considered to be the Granger cause of the target variable [5]. The relationship between the variables is expressed through the following equation:

$$Y_t = \beta_0 + \sum_{i=1}^{m} \beta_i Y_{t-1} + \sum_{i=1}^{m} \alpha_i X_{t-1} + \mu_t \tag{1}$$

$$X_t = \delta_0 + \sum_{i=1}^{m} \delta_i X_{t-1} + \sum_{i=1}^{m} \lambda_i Y_{t-1} + \upsilon_t \tag{2}$$

Where the dependent variable Y represents the spot price of WTI crude oil and X denotes the predictor, in testing the Granger causality between X and Y, the assumptions $\sum_{i=1}^{m} \alpha_i = 0$ , $\sum_{i=1}^{m} \lambda_i = 0$ When none of the original assumptions hold, it means that there is a bi-directional effect between X and Y. When all of the original assumptions hold, it means that X and Y are independent.

The complexity of the crude oil market lies in the complex correlations between the characteristics of its oil price indicators, which makes it difficult to define valid and optimal predictors. Different feature selection methods have different criteria and mechanisms for the validity of the predictors, some of which are quite different.

Based on the above two tests on the crude oil price prediction model screening. First, eight dimensional predictors are integrated as the set of predictors, with WTI crude oil price as the dependent variable. Then, the indicators with causal relationship with crude oil price are screened by

Granger causality test in order to retain the statistically valid predictors while eliminating the indicators with weak performance. This process not only avoids the overfitting problem of the model, but also improves the accuracy of prediction. And the three evaluation criteria, RMSE, $R^2$and MAPE, are used to assess the prediction effect on oil prices, and the research framework is shown in Figure 1.

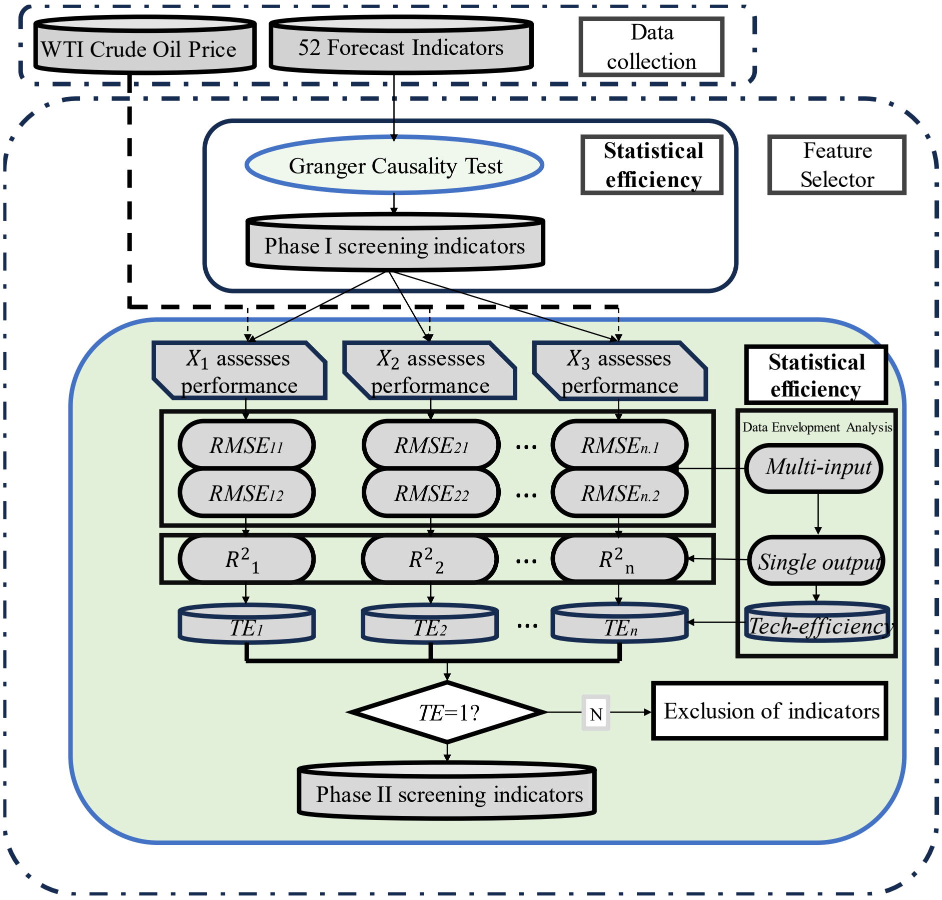


**Figure 1.** Efficiency-based two-stage modeling framework

## 3. Construction of the model

### 3.1. Ridge regression

Ridge regression shows its unique advantages when multiple characteristic variables are involved in the analysis of the crude oil market and there may be some degree of linear dependence between these variables [6]. Its objective function is:

$$L(\beta) = \sum_{i=1}^{n}(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j x_{ij})^2 + \lambda\sum_{j=1}^{p}\beta_j^2 \quad (3)$$

In determining the regularization parameters, this paper combines web search with cross-validation. By systematically traversing multiple $\alpha$values and evaluating the performance of the model under each $\alpha$ value in combination with cross-validation, the optimal $\alpha$value is finally selected. As shown in Figure 2 is the relationship between the mean square error for Alpha and cross-validation

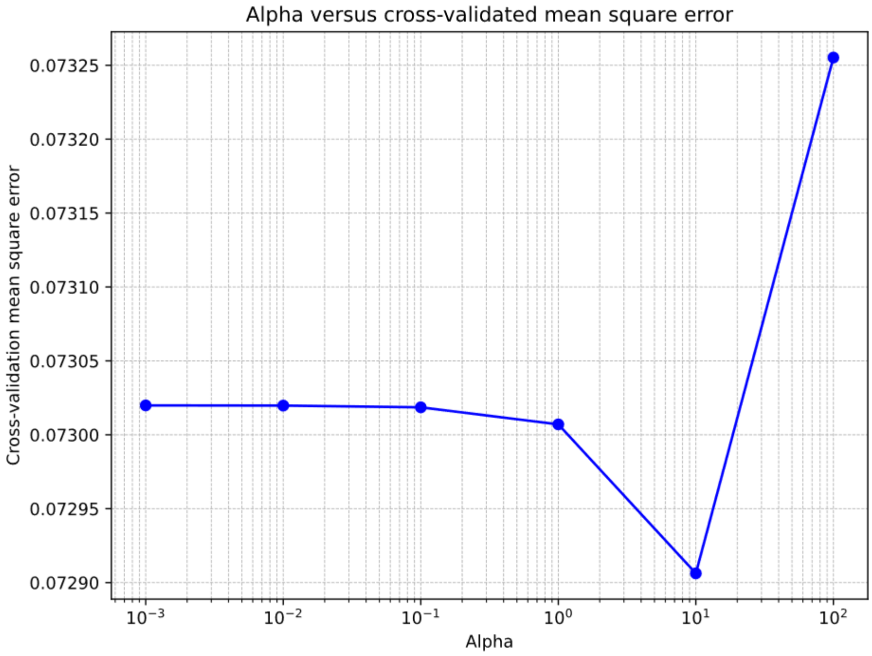


**Figure 2.** Relationship between Alpha and Cross-Validation Mean Square Error

### 3.2. Analysis of experimental results

Artificial Neural Networks (ANN) have been effectively applied to forecast oil prices. ANN demonstrates strong adaptability to the dynamic relationship between diverse and uncertain factors in the crude oil market by employing a nonlinear computational process. A bias term, which is the deviation between the output values and the actual inputs, is also considered in the model as a way to adjust the output of the model [7].

$$u_p = \sum_{i=1}^{n} w_{pi}\, x_i \tag{4}$$

$$y_p = \varphi(u_p + b_p) \tag{5}$$

In this paper, we use multilayer perceptual neural network, three-layer neural network and ReLU activation function is used for activation function.

### 3.3. Random forest

The Random Forest (RF) method effectively models the relationship between multiple predictors and WTI crude oil futures prices [8]. This method provides an accurate prediction of the oil price by calculating the average of the outputs of each decision tree, reducing the risk of overfitting and underfitting. In the algorithm, the decision trees are split based on the selection of variables that minimize the Gini coefficient, which optimizes the prediction accuracy of the model, i. e:

$$I_G(t_{X(x_i)}) = 1 - \sum_{j=1}^{m} f(t_{X(x_i)}, j)^2 \tag{6}$$

$f(t_{X(x_i)}, j)$ denotes the proportion of samples with value $xi$ in leaf j of node t.

### 3.4. Support Vector Regression

In crude oil market analysis, the use of support vector regression (SVR) to deal with complex nonlinear problems and identify high-dimensional features shows its excellent ability [9]. By filtering out valuable support vectors from complex high-dimensional data for computation, the processing difficulty can be significantly reduced, which in turn can predict crude oil prices based on the trained data model. The optimization computational problems involved in implementing this process are as follows:

$$min\frac{1}{2}\|w\|^2 + c\sum_{i=1}^{l}(\xi_i + \xi_i{}^*)\,, s.t. \begin{cases}(y_i - (w^T x_i + b) < \varepsilon + \xi_i \\ (w^T x_i + b) - y_i < \varepsilon + \xi_i{}^* \\ \xi_i, \xi \quad {}^* \geq 0\end{cases} \tag{7}$$

### 3.5. LSTM Neural Networks

The Long Short-Term Memory Network (LSTM) effectively solves the short-term memory problem of traditional neural networks in time series analysis through a complex mechanism of forgetting gates, input gates, and output gates, and strengthens the generalization and fitting ability when facing the price prediction task [10].

The forgetting gate determines which of the previous market information is no longer important. By applying a sigmoid function, the oblivion gate generates a value between 0 and 1 as a way of evaluating how much of each piece of information in the unit state is retained. The oblivion gate discard operation works as follows:

$$f(t) = \sigma(W_{fx}x(t) + W_{fH}h(t-1) + b_f) \tag{8}$$

$$s_c'(t) = s_c(t-1) * f(t) \tag{9}$$

$W_{fx}$, $W_{fH}$ are the weights of the forgetting gate and $b_f$ is its bias.

Input gates are then responsible for deciding what new market information is useful and should be added to the state of the network. For example, if a sudden political event affects the supply of crude oil, the input gate will help the model capture this effect by updating the state of the cell. The formula for input gates to update information is as follows:

$$i(t) = \sigma(W_{iX}x(t) + W_{iH}h(t-1) + b_i) \tag{10}$$

$$g(t) = tanh(W_{gX}x(t) + W_{gH}h(t-1) + b_g) \tag{11}$$

$$s_c(t) = g(t) * i(t) + s_c'(t) \tag{12}$$

$W_{iX}$, $W_{iH}$, $W_{gX}$, $W_{gH}$ are the weights of the input gates and $b_i$ is their bias.

The output gate utilizes the output of t-1 with the input of t. The information of the output is determined by the tanh function. The calculation formula is:

$$o(t) = \sigma(W_{oX}x(t) + W_{oH}h(t-1) + b_o) \tag{13}$$

$$h(t) = tanh(s_c(t) * o(t)) \tag{14}$$

$W_{oX}$, $W_{oH}$ are the weights of the output gates and $b_o$ is their bias.

LSTM was implemented for crude oil price prediction by retaining only the most relevant information for future price movements and adjusting the weights during the training process.

### 3.6. Loss Functions and Performance Metrics

In order to compare the prediction accuracy of five machine learning models, two loss functions with two performance metrics are taken to evaluate the results. The two loss functions are mean square error (MSE) and mean absolute error (MAE), and the performance metrics are R-squared ($R^2$) and adjusted $R^2$ [11]. The loss functions are calculated as:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{15}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \hat{y}_i\right| \tag{16}$$

# 4. Empirical results and analysis

## 4.1. Data description and parameterization

In order to verify whether the results of the feature variable selection in the previous section fit the real driving mechanism of the oil market, this paper uses five different machine learning prediction models for comparison, to verify whether the model predictions are reliable and to derive the best combination of prediction models.

The determination of the optimal α value for the ridge regression algorithm took the method of network search and cross-validation, and was finally set to 10, corresponding to an MSE value size of 0.0729. Considering the large amount of data and the fact that there are much more feature variables than target variables, the ridge regression coefficients use a stochastic mean gradient descent solver. The ratio of training set to prediction set is 7:3.

The multilayer perceptual neural network has the basic structure of a three-layer neural network, and the number of nodes in each layer is determined by the cross-validation method, which is finally set to 100 nodes in each layer; the activation function adopts the ReLU activation function; and the maximum number of iterations is set to 1000. The ratio of training set to prediction set is 7:3.

The random forest algorithm sets the number of decision trees to 100, the number of input variables at each node to be one-half of the total variables, and the ratio of the training set to the prediction set to be 7:3.

The kernel function of the support vector regression algorithm uses a linear kernel function, and the ratio of the training set to the prediction set is 7:3.

The activation function of the LSTM algorithm selects the ReLU function, which solves the situation of gradient disappearance; the number of network layers is selected as 3 layers, and after traversing the algorithm, the optimal number of neurons is 180; the batch size refers to the number of samples used for a single training, which is set to 32 in this paper; regarding the determination of the number of iterations, the first iteration is set to be 1,000, to observe the trend of the mean-square error MSE with the number of iterations, which is shown in the following figure. as shown in the figure 3 below:

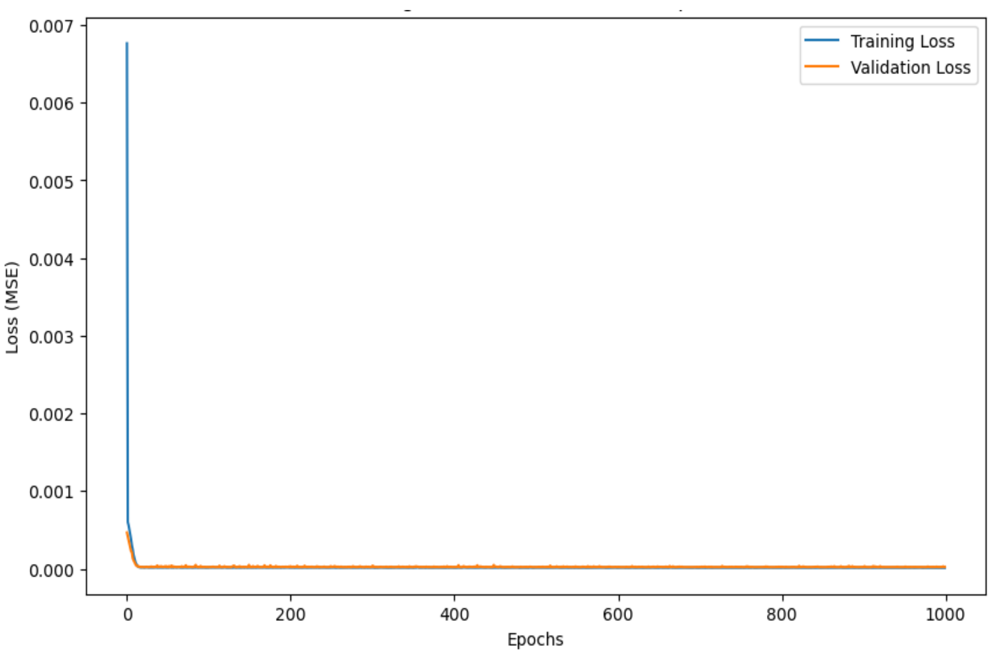


**Figure 3.** Plot of mean square error MSE with number of iterations

It is observed that the error between the validation set and the prediction set has leveled off after 500 times, so the number of iterations is finally set to 600. The ratio of training set to prediction set is 7:3.

### 4.2. Assessment of model prediction performance

This study used the GCT-RFE model combined with five different machine learning prediction algorithms to compute model performance under two different loss functions and two performance metrics criteria. Specifically, the mean square error (MSE), mean absolute error (MAE), $R^2$, and adjusted $R^2$ of each combined model were evaluated. Then, the prediction performance of each combined model was comprehensively evaluated. The comparative results of the model prediction performance assessment are detailed in Table 1.

**Table 1.** Comparative results of the assessment of the predictive performance of the models

| Model | Localization | MSE | MAE | $R^2$ | Adjusted $R^2$ |
|---|---|---|---|---|---|
| Ridge regression | Training set | 0.566 | 0.436 | 0.9745 | 0.9745 |
| | Forecast set | 0.914 | 0.493 | 0.9516 | 0.9513 |
| ANN | Training set | 0.423 | 0.374 | 0.9732 | 0.9731 |
| | Forecast set | 0.613 | 0.452 | 0.9608 | 0.9607 |
| RF | Training set | 0.769 | 0.513 | 0.9641 | 0.9639 |
| | Forecast set | 0.966 | 0.736 | 0.9124 | 0.9119 |
| SVR | Training set | 0.771 | 0.570 | 0.9744 | 0.9743 |
| | Forecast set | 0.822 | 0.625 | 0.9512 | 0.9509 |
| LSTM | Training set | 0.589 | 0.945 | 0.9754 | 0.9753 |
| | Forecast set | 0.975 | 1.075 | 0.9354 | 0.9349 |

Combined with the table, it is easy to see that among the five machine learning prediction algorithms, the artificial neural network stands out, with its prediction set having the smallest MSE value and the largest adjusted $R^2$ value, in which the adjusted $R^2$ value reaches 0.9607. At the same time, the difference between the MSE and MAE values of the training set and prediction set is extremely small, which excludes the case of model overfitting.

### 4.3. Robustness Tests

Robustness testing is a critical step to ensure that the model can remain reliable and valid despite global political and economic fluctuations and even unexpected events such as economic crises or geopolitical conflicts.

#### 4.3.1. Generating Adversarial Samples

The prediction accuracy of the model on unseen data can also be improved by training the model to recognize and correctly process adversarial samples [12]. The confrontation sample generation formula can be expressed as:

$$\boldsymbol{X_{adv}} = \boldsymbol{X} + \boldsymbol{\varepsilon} \cdot \boldsymbol{N}(\mathbf{0}, \mathbf{1}) \tag{17}$$

$X_{adv}$ is the generated adversarial sample, $X$ is the original input feature matrix, $\varepsilon$ is the perturbation strength, and $N(1,0)$ is a random noise of the same $X$ and shape generated from a standard normal distribution.

#### 4.3.2. Empirical evidence for robustness testing

The paper sets the perturbation strength $\boldsymbol{\varepsilon}$ to 0.01, indicating that the perturbation performed on each feature value is 1% of its value. In order to verify whether the five machine learning prediction models in the previous paper are robust, by generating adversarial samples in the original dataset, the magnitude of the MSE values, MAE values and adjusted $R^2$ values of the original prediction set compared with those of the generated adversarial samples and the test set are compared, and the comparison results are shown in Table 2. Post-disturbance prediction set (PPS).

**Table 2.** Comparison of model robustness metrics

| Model | Localization | MSE | MAE | Adjusted $R^2$ |
|---|---|---|---|---|
| Ridge regression | Forecast set | 0.914 | 0.493 | 0.9516 |
| | PPS | 0.918 | 0.534 | 0.9514 |
| ANN | Forecast set | 0.423 | 0.374 | 0. 9607 |
| | PPS | 0.435 | 0.352 | 0.9605 |
| RF | Forecast set | 0.966 | 0.736 | 0.9124 |
| | PPS | 1.121 | 0.858 | 0.9256 |
| SVR | Forecast set | 0.822 | 0.625 | 0.9512 |
| | PPS | 0.919 | 0.648 | 0.9511 |
| LSTM | Forecast set | 0.975 | 1.075 | 0.9354 |
| | PPS | 4.842 | 5.248 | 0.6790 |

According to the results of the robustness test, the ridge regression model, artificial neural network model, random forest model, and support vector regression model do not produce much interference in the model accuracy after fighting against the sample interference, and the artificial neural network is still the optimal model, which shows that the combination of the processed data of the ADF and the GCT model and the artificial neural network algorithm can realize the effective prediction of the WTI crude oil futures price in the market and can withstand the volatility of the data.

## 5. Conclusion and outlook

### 5.1. Conclusions of the study

The future direction of international crude oil prices is a complex and volatile issue, which is influenced by multiple factors such as global economic conditions, supply and demand, geopolitics, monetary policy, and new energy development. This paper proposes a combined model of crude oil market driver screening and price system forecasting, aiming to gain a deeper understanding of the dynamic factors and forecasting mechanisms of international crude oil prices. It is finally concluded that the artificial neural network based on ADF and GCT model processing has the best prediction performance, and the model has good robustness, which can make effective combing and prediction of the crude oil market, and provide an important reference for policy makers to adjust the energy strategy.

### 5.2. Insights and recommendations

Through the findings of this paper, the following insights can be brought about:

The future direction of international crude oil prices will be influenced and constrained by multiple factors. In the short term, oil prices may be volatile due to geopolitical tensions, global economic recovery and monetary policy adjustments; in the medium to long term, they may be downward trending due to the development of new energy sources, energy transformation and global economic restructuring. Therefore, investors and policymakers will need to pay close attention to market dynamics and risk factors in the future in order to formulate reasonable investment strategies and policy measures.

## References

[1] Zhao W L, Fan Y, Ji Q. Extreme risk spillover between crude oil price and financial factors [J]. Finance Research Letters, 2022, 46: 102317.

[2] Li M, Cheng Z, Lin W, et al. What can be learned from the historical trend of crude oil prices? An ensemble approach for crude oil price forecasting[J]. Energy Economics, 2023, 123: 106736.

[3] Gupta N, Nigam S. Crude oil price prediction using artificial neural network [J]. Procedia Computer Science, 2020, 170: 642 - 647.

[4] Orzeszko W. Nonlinear causality between crude oil prices and exchange rates: Evidence and forecasting [J]. Energies, 2021, 14 (19): 6043.

[5] Shojaie A, Fox E B. Granger causality: A review and recent advance s [J]. Annual Review of Statistics and Its Application, 2022, 9 (1): 289 - 319.

[6] Hoerl R W. Ridge regression: a historical context [J]. Technometrics, 2020, 62 (4): 420 - 425.

[7] Abdolrasol M G M, Hussain S M S, Ustun T S, et al. Artificial neural networks-based optimization techniques: A review [J]. Electronics, 2021, 10 (21): 2689.

[8] Nyangarika A, Mikhaylov A, Muyeen S M, et al. Energy stability and decarbonization in develo** countries: Random Forest approach for forecasting of crude oil trade flows and macro indicators [J]. Frontiers in Environmental Science, 2022, 10: 1031343.

[9] Zhang F, O'Donnell L J. Support vector regression[M]//Machine learning. Academic Press, 2020: 123 - 140.

[10] Zhang K, Hong M, Zhang K, et al. Forecasting crude oil price using LSTM neural networks [J]. Data Sci. Financ. Econ, 2022, 2: 163 - 180.

[11] Wang Q, Ma Y, Zhao K, et al. A comprehensive survey of loss functions in machine learning [J]. Annals of Data Science, 2020: 1 - 26.

[12] Qian Z, Huang K, Wang Q F, et al. A survey of robust adversarial training in pattern recognition: Fundamental, theory, and methodologies [J]. Pattern Recognition, 2022, 131: 108889.